UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/421,416	10/19/1999	HINRICH SCHUETZE	D/99198	4124

7590 08/19/2002

JOHN E BECK
XEROX CORPORATION
XEROX SQUARE 20A
ROCHESTER, NY 14644

EXAMINER	
FLEURANTIN, JEAN B	

ART UNIT	PAPER NUMBER
2172	8

DATE MAILED: 08/19/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/421,416	SCHUETZE ET AL.	
	Examiner	Art Unit	
	Jean B Fleurantin	2172	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *14 June 2002* .

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,7-39 and 41-49* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,7-22,28,39,41-45,47 and 49* is/are rejected.

7)☒ Claim(s) *23-27,46,48 and 2938* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) 5 .
4) ☐ Interview Summary (PTO-413) Paper No(s). _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: .

DETAILED ACTION

1. Claims 2-6 and 40 are canceled.

Claims 43-49 are added.

Claims 1, 7-39 and 41-49 are remained pending for examination.

Response to Amendment

2. Applicant's arguments submitted on 06/14/2002 with respect to claims 1, 7-39 and 41-49 have been considered but are not persuasive. Examiner discusses the new added claims 43-49 in the following rejection.

Response to Applicant' Remarks

3. On page 15, Applicant stated that 'Schuetze does not teach using a feature related images included in the documents.' However, Examiner disagrees because Li includes the steps of the anchor text may also be in the form of images graphics, etc. so the index engine may substitute other information such as the tail document's title for the non-textual anchor text; which is readable as first feature comprising text surrounding an image included in the document (see col. 10, lines 49-52). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Schuetze and Li with the step of text surrounding an image included in the document. Also, in column 1, lines 12 through 16, Li further teaches steps of non-sequential method of accessing information using nodes and links nodes, i.e. documents or files, contain text graphics audio video animation and images while links connect the nodes or documents to other nodes or documents. This modification would

allow the teachings of Schuetze and Li to improve the accuracy of the system and method for quantitatively representing data objects in vector space, and provide comparison the words in the query to the words in a hyperlink to obtain a relevance ranking for each hyperlink and summing the relevance rankings for each hyperlink pointing to a particular document to obtain a summed relevance score for that document (see col. 4, lines 19-24).

On page 16, Applicant stated that 'Schuetze does not teach or suggest all the claim limitations.' Although, Schuetze does not explicitly teach all the claim limitations it teaches the system in the art; see col. 4, lines 9-16.

In response to applicant's argument on pages 16 and 17, that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

Examiner is entitled to give claim limitations their broadest reasonable interpretation in light of the specification.

Interpretation of Claims-Broadest Reasonable Interpretation

During patent examination, the pending claims must be 'given the broadest reasonable interpretation consistent with the specification.' Applicant always has the opportunity to amend

the claims during prosecussion and broad interpretation by the examiner reduces the possibility that the claim, once issued, will be interpreted more broadly than is justified. In re Prater, 162 USPQ 541,550-51 (CCPA 1969).

Claim Rejections - 35 U.S.C. § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1, 7-22, 28, 39, 41-45, 47 and 49 are rejected under 35 U.S.C. 103(a) as being unpatentable over Schuetze (US Pat. 5,675,819) in view of Li (US Pat. 5,920,859), ("Schuetze"), ("Li").

As per claims 1 and 39, Schuetze substantially teaches a method for quantitatively representing objects in a vector space, as claimed comprises the steps of identifying a first document to be processed from a plurality of objects documents (thus, a search is performed to retrieve possibly relevant documents the documents are analyzed to determine the number that are actually relevant to the query, the precision of the search is the ratio of the number of relevant documents to the number of retrieved documents; which is readable as identifying a first document to be processed from a plurality of objects documents) (see col. 18, lines 63-67);

converting the first feature to a first vector (thus, in computing a document vector, those terms that correspond to the sense used in the document will be reinforced whereas the direction

represented by the inappropriate sense will not be present in other words, which is readable as converting the first feature to a first vector) (see col. 8, lines 14-18);

associating the first vector with the first document (thus, each term of the documents is associated with a vector that represents the term's pattern of local co-occurrences, this vector can then be compared with others to measure the co-occurrence similarity, and hence semantic similarity of terms; which is readable as associating the first vector with the first document) (see col. 6, lines 27-32); also in column 5, lines 4 through 10, Schuetze further teaches after forming the thesaurus vectors, a context vector for each document is computed, the context vector is a combination of the weighted sums of the thesaurus vectors of all the words contained in the document, these context vectors then induce a similarity measure on documents and queries that can be directly compared to standard vector-space methods;

extracting a first feature corresponding to the first document from the plurality of documents(thus, accessing and browsing documents based on content similarity, words and documents are represented as vectors in the same multi-dimensional space that is derived from global lexical co-occurrence patterns; which is readable as extracting a first feature corresponding to the first document from the plurality of documents) (see col. 4, lines 9-13). But, Schuetze does not explicitly indicate the steps of the first feature comprising text surrounding an image included in the document. However, Li implicitly teaches the step of anchor text may also be in the form of images graphics, etc. so the index engine may substitute other information such as the tail document's title for the non-textual anchor text; which is

readable as first feature comprising text surrounding an image included in the document (see col. 10, lines 49-52). Also, in column 1, lines 12 through 19, Li further teaches steps of non-sequential method of accessing information using nodes and links nodes, i.e. documents or files, contain text graphics audio video animation and images while links connect the nodes or documents to other nodes or documents, the most popular hypertext or hypermedia system is the World Wide Web, which links various nodes or documents together using hyperlinks. Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Schuetze and Li with the step of text surrounding an image included in the document. This modification would allow the teachings of Schuetze and Li to improve the accuracy and the reliability of the system and method for quantitatively representing data objects in vector space, and provide comparison the words in the query to the words in a hyperlink to obtain a relevance ranking for each hyperlink and summing the relevance rankings for each hyperlink pointing to a particular document to obtain a summed relevance score for that document (see col. 4, lines 19-24).

As per claim 7, Schuetze substantially teaches a method as claimed, further comprises the steps of converting the second feature to a second vector (thus, context vectors then introduce a similarity measure on documents and queries that can be directly compared to standard vector space methods; which is readable as converting the second feature to a second vector) (see col. 5, lines 7-10); and

associating the second vector with the first document (thus, a context vector for each document is computed, which is readable as converting the second feature to a second vector) (see col. 5, lines 4-5). But, Schuetze does not explicitly indicate the steps of extracting a second feature corresponding to the document, the second feature comprising a first URL representing the first documents. However, Li implicitly teaches the step of the query may be represented by a query vector where the query vector contains a dimension for each term in the query, each document may be represented by document link vectors for each hyperlink pointing to the document, where each document link vector contains a dimension for each term in the corresponding hyperlink pointing to that document comparing the words in the query to the words in the hyperlinks includes calculating the dot product of the query vector with the document link vector for that hyperlink summing the relevance ranking for each hyperlink pointing to a document includes summing the dot products obtained using the document link vectors for a particular document to obtain the summed relevance score for that document, the summed relevance scores may then be compared to obtain a ranking of documents; which is readable as URL representing first documents (see col. 4, lines 25-39). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Schuetze and Li with the step of URL representing first documents. This modification would allow the teachings of Schuetze and Li to improve the accuracy and the reliability of the system and method for quantitatively representing data objects in vector space, and provide comparison the words in the query to the words in a hyperlink to obtain a relevance

ranking for each hyperlink and summing the relevance rankings for each hyperlink pointing to a particular document to obtain a summed relevance score for that document (see col. 4, lines 19-24).

As per claims 8, 11, 13 and 18-19, the limitations of claims 8, 11, 13 and 18-19 are rejected in the analysis of claim 7 above, and these are rejected on that basis.

As per claims 9 and 14, the limitations of claims 9 and 14 are rejected in the analysis of claim 7 above, and these are rejected on that basis.

As per claims 10 and 15, in addition to the discussion in claim 7 above, Schuetze teaches all the subject matter of the claimed invention with the exception of an exact the second feature comprising inlinks in the collection of documents linking to the first document; and the second feature comprising outlinks in the collection of documents linking of the first document. However, Li teaches the steps of each document may be represented by document link vectors for each hyperlink pointing to the document, where each document link vector contains a dimension for each term in the corresponding hyperlink pointing to that document comparing the words in the query to the words in the hyperlinks includes calculating the dot product of the query vector with the document link vector for that hyperlink; which is readable as inlinks in the collection of documents linking to the first document; and the second feature comprising outlinks in the collection of documents linking of the first document(see col. 4, lines 25-33). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Schuetze and Li with the step of inlinks in the collection of documents

linking to the first document; and the second feature comprising outlinks in the collection of documents linking of the first document. This modification would allow the teachings of Schuetze and Li to improve the accuracy and the reliability of the system and method for quantitatively representing data objects in vector space.

As per claim 12, Schuetze substantially teaches a method as claimed, wherein the numeric value representative of the number of links in each corresponding document linking to the document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link (thus, two documents are considered similar if they share a significant number of terms with medium frequency terms preferentially weighted terms are then grouped by their occurrence in these document clusters, since a complete-link document clustering is performed, the procedure is very computationally intensive and does not scale to a large reference corpus, further the central assumption that terms are related if they often occur in the same documents seems problematic for corpora with long documents; which is readable as wherein the numeric value representative of the number of links in each corresponding document linking to the document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link) (see col. 2, lines 51-60).

As per claims 16 and 41, in addition to the discussion in claim above, Schuetze teaches the step of counting the occurrences of each unique word in the subject document (thus, the dimensionality of the thesaurus space is reduced by using a singular value decomposition the

closeness of terms with equal frequency occurs because the terms have about the same number of zero entries in their term vectors, for a given term singular value decomposition assigns values to all dimensions of the space, so that frequent and infrequent terms can be close in the reduced space if they occur with similar terms, for example, the word "accident," which may occur 2590 times, and the word "mishaps," which may occur only 129 times, can have similar vectors that are close despite the frequency difference between them, the technique of singular value decomposition (SVD) is used to achieve a dimensional reduction by obtaining a compact and tractable representation for search purposes, the uniform representation for words and documents provides a simple and elegant user interface for query focusing and expansion; which is readable as counting the occurrences of each unique word in the subject document) (see cols. 4-5, lines 54-3);

creating a vector having a number of dimensions equal to the number of unique words in the collection of documents, and further having as each element a numeric value representative of the number of occurrences in the subject document of the corresponding word (thus, terms are represented as high-dimensional vectors with a component for each document in the corpus, the value of each component is a function of the frequency the term has in that document they show that query expansion using the cosine similarity measure on these vectors improves retrieval performance; however, the time complexity for computing the similarity between terms is related to the size of the corpus because the term vectors are high-dimensional (see col. 3, lines 8-17).

As per claims 17 and 42, Schuetze substantially teaches a method as claimed, wherein the value representative of the number of occurrences in the subject document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word (thus, documents are clustered into small groups based on similarity measure two documents are considered similar if they share a significant number of terms with medium frequency terms preferentially weighted terms are then grouped by their occurrence in these document clusters, since a complete-link document clustering is performed the procedure is very computationally intensive and does not scale to a large reference corpus; which is readable as wherein the value representative of the number of occurrences in the subject document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word (see col. 2, lines 51-57). Also, in column 17, lines 30 through 44, Schuetze teaches the step of weighting the words in the document is by using an augmented tf.idf method 'term frequency-inverse document frequency method' when summing thesaurus vectors: ##EQU13## where tf.sub.ij is the frequency of word I in document j; N is the total number of documents; and n.sub.i is the document frequency of word I. as the word frequency increases in a document, the weight (score) for that word also increases, however, the term N/n.sub.i is inversely proportional to document frequency such that high frequency words receive less weight.

As per claim 20, Schuetze substantially teaches a method as claimed, wherein the feature comprises text represented by the subject document (thus, methods perform a computation on the text of the documents in the corpus to produce a thesaurus, which is readable as text represented by the subject document) (see col. 2, lines 17-18). Also, in column 1, lines 14 through 20, Schuetze teaches the step of the information retrieval systems typically define similarity between queries and documents in terms of a weighted sum of matching words, the usual approach is to represent documents and queries as long vectors and use similarity search techniques.

As per claim 21, in addition to the discussion in claim 5 above, Schuetze teaches the step of wherein the converting step comprises the steps of for each possible text genre, processing the subject document to calculate the probability that the subject document is of the corresponding genre (thus, the documents are analyzed to determine the number that are actually relevant to the query, the precision of the search is the ration of the number of relevant documents to the number of retrieved documents; which is readable as processing the subject document to calculate the probability that the subject document is of the corresponding genre) (see col. 18, lines 64-67).

As per claims 22, 28, 45 and 47, the limitations of claims 22, 28, 45 and 47 are rejected in the analysis of claim 1 above, and these are rejected on that basis.

As per claim 43, in addition to the discussion in claim 1 above, Schuetze teaches converting information associated with the second feature into a second vector (thus, context vectors then introduce a similarity measure on documents and queries that can be directly

compared to standard vector space methods; which is readable as converting information associated with the second feature into a second vector) (see col. 5, lines 7-10); and

associating the second vector with the document (thus, a context vector for each document is computed, which is readable as associating the second vector with the document) (see col. 5, lines 4-5)

As per claim 44, the limitations of claim 44 are rejected in the analysis of claim 43 above, and this is rejected on that basis.

As per claim 49, in addition to the discussion in claims 1 and 43 above, Schuetze teaches all the subject matter of the claimed invention with the exception of an exact the second feature comprising a one of a text feature, a hyperlink feature, a user feature and a genre feature. However, Li teaches the steps of each document may be represented by document link vectors for each hyperlink pointing to the document, where each document link vector contains a dimension for each term in the corresponding hyperlink pointing to that document comparing the words in the query to the words in the hyperlinks includes calculating the dot product of the query vector with the document link vector for that hyperlink; which is readable as exact the second feature comprising a one of a text feature, a hyperlink feature, a user feature and a genre feature (see col. 4, lines 25-33). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Schuetze and Li with the step of the second feature comprising a one of a text feature, a hyperlink feature, a user feature and a genre feature. This modification would allow the teachings of Schuetze and Li to improve the accuracy

and the reliability of the system and method for quantitatively representing data objects in vector space, and provide a unique and non sequential method of accessing information using nodes and links (see col. 1, lines 11-13).

Allowable Subject Matter

5. Claims 23-27, 29-38, 46 and 48 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

6. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure: Rose at al. US Pat. No. 5,870,740 relates to an information retrieval system. Corey et al. US Pat. No. 5,987,446 relates to text searching engine are utilized in searching for one or more desired information items. Deerwester US Pat. No. 5,778,362 relates to methods and systems for analyzing collections of data items to reveal structures such as associative structures within the collections of data items. Bolle et al. US Pat. No. 5,546,475 relates to the field of recognizing.

7. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL.** See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after

the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Conclusion

8. Any inquiry concerning this communication from examiner should be directed to Jean Bolte Fleurantin at (703) 308-6718. The examiner can normally be reached on Monday through Friday from 7:30 A.M. to 6:00 P.M.

If any attempt to reach the examiner by telephone is unsuccessful, the examiner's supervisor, Mrs. KIM VU can be reached at **(703) 305-8449.** The FAX phone numbers for the Group 2100 Customer Service Center are: ***After Final* (703) 746-7238**, ***Official* (703) 746-7239**, and ***Non-Official (703) 746-7240***. NOTE: Documents transmitted by facsimile will be entered as official documents on the file wrapper unless clearly marked "***DRAFT***".

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Group 2100 Customer Service Center receptionist whose telephone numbers are **(703) 306-5631, (703) 306-5632, (703) 306-5633**.

Jean Bolte Fleurantin

August 12, 2002

JBF/

JEAN M. CORRIELUS
PRIMARY EXAMINER